



**XAI FOR WINE QUALITY CLASSIFICATION USING SHAPLEY ADDITIVE
EXPLANATIONS IN INTERPRET FEATURE IMPORTANCE IN SUPERVISED
LEARNING**

FINAL THESIS

IHSANUL ARIFIN RAMADHANI

41522010249

**INFORMATICS ENGINEERING STUDY PROGRAM
FACULTY OF COMPUTER SCIENCE
INTERNATIONAL UNDERGRADUATE PROGRAM
MERCU BUANA UNIVERSITY**

JAKARTA

2026



**XAI FOR WINE QUALITY CLASSIFICATION USING SHAPLEY ADDITIVE
EXPLANATIONS IN INTERPRET FEATURE IMPORTANCE IN SUPERVISED
LEARNING**

FINAL THESIS

**IHSANUL ARIFIN RAMADHANI
41522010249**

**SUBMITTED AS ONE OF THE REQUIREMENTS FOR
PRESENTATION AT THE FINAL THESIS DEFENSE**

INFORMATICS ENGINEERING STUDY PROGRAM

FACULTY OF COMPUTER SCIENCE

INTERNATIONAL UNDERGRADUATE PROGRAM

MERCU BUANA UNIVERSITY

JAKARTA

2026

PAGE OF STATEMENT OF ORIGINAL WORK

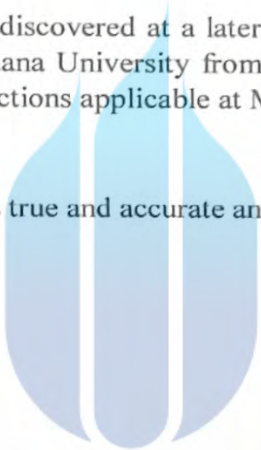
I, the undersigned below :
Name : Ihsanul Arifin Ramadhani
Student Number (NIM) : 41522010249
Study Program : Informatics Engineering

Declares truthfully that the final project entitled:

“Xai For Wine Quality Classification Using Shapley Additive Explanations in Interpret Feature Importance in Supervised Learning” this is my own work, does not contain any elements of plagiarism, copyright infringement, or illegal content in any form, and does not violate the laws or rights of any party.

If any violation of this statement is discovered at a later date, I am willing to bear all legal consequences and release Mercu Buana University from any form of legal claim, and I am prepared to receive the academic sanctions applicable at Mercu Buana University.

I hereby declare that this statement is true and accurate and is to be used as appropriate.


J. 10th 2026
10000
METERAL
TEMPEL
7B489ANX195390390
Ihsanul Arifin Ramadhani
UNIVERSITAS
MERCU BUANA

PERNYATAAN SIMILARITY CHECK
/SIMILARITY CHECK STATEMENT

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh
/The undersigned, hereby declare that the scientific work written by

Nama /Name : Ihsanul Arifin Ramadhani
NIM /Student ID Number : 41522010249
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

/The title:

“XAI FOR WINE QUALITY CLASSIFICATION USING SHAPLEY ADDITIVE EXPLANATIONS IN INTERPRET FEATURE IMPORTANCE IN SUPERVISED LEARNING”

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

/Similarity checks have been carried out with the Turnitin system on the date:

10 Januari 2026

dengan nilai persentase sebesar :

/with a percentage value of:

9%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer** Universitas Mercu Buana. */declared to meet standards in accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

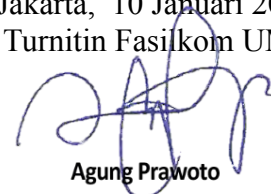
File hasil cek similarity turnitin:

/Turnitin similarity report file

https://drive.google.com/file/d/1Lqu3cy1wr7TZFx5SfjQkfi_j80som-m3/view?usp=drive_link



Jakarta, 10 Januari 2026
Admin Turnitin Fasilkom UMB



Agung Prawoto, S.Kom., B.Sc
NIK : 322970503

ENDORSEMENT PAGE

This Final Project is submitted by :

Name : Ihsanul Arifin Ramadhani
Student Number (NIM) : 41522010249
Study Program : Informatics Engineering
Title : Xai For Wine Quality Classification Using Shapley
Additive Explanations in Interpret Feature Importance in
Supervised Learning

Has been successfully defended in the examination on January 12, 2026, and accepted as part of the requirements to obtain a Bachelor's degree in the Informatics Engineering Study Program, Faculty of Computer Science, Universitas Mercu Buana.

Approved by:

Supervisor



Ilham Nugraha, S.Kom., M.Sc.

NIDN/NUPTK: 0307098904

UNIVERSITAS

Jakarta, 12 January 2026

Acknowledged by,

MERCU BUANA

Dean of the Faculty of Computer Science

Head of Informatics Engineering
Study Program



Dr. Bambang Jokonowo, S.Si., MTI

NIDN/NUPTK: 0320037002



Dr. Hadi Santoso, S.Kom., M.Kom

NIDN/NUPTK: 0225067701

MOTTO & ACKNOWLEDGEMENTS

Whatever progress I have made rests upon the quiet sacrifices of my closest people. Their steadfast resilience and unspoken support have shaped my path, and this thesis stands as a testament to the values they have lived, perseverance, humility, and unwavering devotion.

First and foremost:

1. To my beloved parents Mum and Dad, Rina Sumarni and Yusya Januar, I dedicate this thesis to you. You carried what I did not see, endured what you rarely spoke of, and kept going when it would have been easier to stop. Your love was not loud, but it was constant, not performative, but steadfast. If there is any strength in these pages, it is because I learned endurance from you quietly, over time, through the way you lived.
2. To my partner, Ditha Firsty Auguestien, thank you for being my steadier horizon throughout this journey. For the patience in the long seasons of work, the kindness in moments of doubt, and the gentle reminders to keep going when I felt spent, your support has mattered more than you know. This thesis carries, between its lines, the calm you gave me when I needed it most.
3. To the lecturers of the International Undergraduate Informatics Engineering Programme at Universitas Mercu Buana, I extend my sincere appreciation. Thank you for the discipline you upheld and the knowledge you entrusted to your students. You shaped not only what I know, but how I think, how I question, refine, and pursue clarity with rigour.
4. I would also like to express my sincere gratitude to my supervisor, Ilham Nugraha, S.Kom., M.Sc., for his guidance, constructive feedback, and steady support throughout the thesis process. His suggestions helped sharpen both the direction of the research and the clarity of its presentation, and his supervision made this work far more structured and achievable.

5. To my coursemates four companions from the same cohort and programme Joao Nitoivar Ribeiro Cancio, David Ginola Guterres, Muhammad Ilham Haekal, and M. Chandra Agoeng Pramudya Abdullah thank you for the camaraderie that carried me through ordinary weeks and difficult deadlines alike.
6. My thanks also go to my friends from my daily circle Rais Khan, Muhammad Mahrus Irsyam, and Hamzah Nuriadi for the encouragement, humour, and companionship that kept me grounded.
7. I am also grateful to the places that became part of my routine and, in their own way, part of this thesis. Thank you to Madurani, which at times offered a reliable corner for drafting and revising, and to Bu'de and the warmth of that small shop environment that made working there feel welcoming. My appreciation also extends to Kopi Kenangan, which often became my workspace for revisions and long writing sessions.
8. And, on a lighter note, I would like to thank the craft of coffee itself especially espresso, Americano, latte, and aren latte which became reliable fuel through long hours of analysis, coding, and writing.
9. Last but not least, I wanna thank me for believing in me, I wanna thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting. I wanna thank me for always being a giver and trying to give more than I receive. I wanna thank me for trying to do more right than wrong. I wanna thank me for being me at all times.

ABSTRACT

Name : Ihsanul Arifin Ramadhani
Student ID : 41522010249
Study Program : Undergraduate Informatics Engineering Programme
Title of Research : *XAI for Wine Quality Classification Using Shapley Additive Explanations in Interpreting Feature Importance in Supervised Learning*
Supervisor : Ilham Nugraha, S.Kom., M.Sc.

This study investigates explainable supervised learning for red wine quality classification using the UCI Wine Quality (red wine) dataset (1,599 samples; 11 physicochemical features). The ordinal quality score is converted into a binary label (quality ≤ 5 : low; quality ≥ 6 : high). Using an 80/20 stratified split, Logistic Regression, Random Forest, and XGBoost are compared, with F1-score as the primary metric. XGBoost delivers the best held-out test performance, achieving $F1 = 0.8452$ at the default 0.50 threshold. A single global decision threshold is then selected from out-of-fold probabilities via stratified cross-validation on the training set, avoiding test-set tuning. The optimal threshold of 0.45 yields a modest improvement on the test set ($F1 = 0.8487$) with slightly higher recall. SHAP explanations show that alcohol and sulphates typically raise high-quality likelihood, while volatile acidity and total sulphur dioxide push predictions towards low quality. These findings support accurate yet transparent quality screening.

Keywords: Explainable AI, SHAP, XGBoost, supervised learning, wine quality classification, threshold calibration, feature importance.

FINAL PROJECT PUBLICATION APPROVAL STATEMENT PAGE FOR THE UNIVERSITY REPOSITORY

As an academic member of Universitas Mercu Buana, I, the undersigned:

Name : Ihsanul Arifin Ramadhani
Student Number (NIM) : 41522010249
Study Program : Informatics Engineering
Title : Xai For Wine Quality Classification Using Shapley
Additive Explanations in Interpret Feature
Importance in Supervised Learning

For the purpose of the development of science and knowledge, I hereby grant permission and agree to provide Universitas Mercu Buana with a **Non-exclusive Royalty-Free Right** for my scientific work titled above, including any accompanying materials (if required).

Under this Non-exclusive Royalty-Free Right, Universitas Mercu Buana reserves the right to store, transfer media/format, manage within a database, maintain, and publish my Final Project, provided that my name is always cited as the author/creator and as the owner of the Copyright.

I make this statement in truth.

UNIVERSITAS
MERCU BUANA

January 12th January 2026

METERAI
TEMPEL
E5ANX195390379
Ihsanul Arifin Ramadhani

ABSTRAK

Nama : Ihsanul Arifin Ramadhani
NIM : 41522010249
Program Studi : Undergraduate Informatics Engineering Programme
Judul Penelitian : *XAI for Wine Quality Classification Using Shapley Additive Explanations in Interpreting Feature Importance in Supervised Learning*
Pembimbing : Ilham Nugraha, S.Kom., M.Sc.

Studi ini meneliti penerapan *explainable supervised learning* untuk klasifikasi kualitas red wine menggunakan dataset UCI Wine Quality (red wine) (1.599 sampel; 11 fitur fisikokimia). Skor kualitas ordinal dikonversi menjadi label biner (kualitas ≤ 5 : rendah; kualitas ≥ 6 : tinggi). Dengan *stratified split* 80/20, Logistic Regression, Random Forest, dan XGBoost dibandingkan, dengan F1-score sebagai metrik utama. XGBoost memberikan kinerja terbaik pada *held-out test set*, mencapai $F1 = 0,8452$ pada ambang (threshold) default 0,50. Selanjutnya, satu ambang keputusan global dipilih dari probabilitas *out-of-fold* melalui *stratified cross-validation* pada data latih, sehingga menghindari penyesuaian ambang pada data uji. Ambang optimal 0,45 menghasilkan peningkatan kecil pada data uji ($F1 = 0,8487$) dengan recall yang sedikit lebih tinggi. Penjelasan SHAP menunjukkan bahwa alkohol dan sulphates umumnya meningkatkan peluang kualitas tinggi, sedangkan *volatile acidity* dan *total sulphur dioxide* mendorong prediksi ke kualitas rendah. Temuan ini mendukung penyaringan kualitas yang akurat sekaligus transparan.

Kata Kunci: Explainable AI, SHAP, XGBoost, supervised learning, wine quality classification, threshold calibration, feature importance.

TABLE OF CONTENTS

COVER	0
TITLE.....	i
PAGE OF STATEMENT OF ORIGINAL WORK	ii
TURNITIN TEST RESULT STATEMENT PAGE.....	iii
ENDORSEMENT PAGE	iv
MOTTO & ACKNOWLEDGEMENTS	v
ABSTRACT	vii
FINAL PROJECT PUBLICATION APPROVAL STATEMENT PAGE FOR THE UNIVERSITY REPOSITORY	viii
ABSTRAK.....	ix
TABLE OF CONTENTS	x
TABLE OF FIGURES.....	xiii
TABLES	xiv
CHAPTER I INTRODUCTION.....	1
1.1 Background	1
1.1 Problem Statement.....	3
1.2 Research Objectives.....	3
1.4 Scope and Limitations.....	4
1.5 Significance of the Study.....	5
CHAPTER II LITERATURE REVIEW	6
2.1 Main Theory.....	6
2.1.1 Machine Learning and Supervised Learning	6
2.1.2 Classification in Machine Learning.....	6
2.1.3 Wine Dataset and Physicochemical Attributes	7
2.1.4 Explainable Artificial Intelligence (XAI)	7
2.1.5 Shapley Value & SHAP	8
2.1.6 Classification Model Evaluation	9
2.2 Supporting Theory.....	9
2.2.1 Decision Tree and Random Forest Algorithms	9
2.2.2 Data Preprocessing and Class Imbalance Handling	10

2.2.3 Visualization and Interpretation of SHAP	11
2.2.4 Wine Quality Domain	12
2.3 Preliminary Studies	12
CHAPTER III METHODOLOGY	17
3.1 Research Design	17
3.2 Data Source	18
3.3 Data Description	19
3.3.1 Physicochemical Attributes	20
3.4 Statistical and Mathematical Framework	21
3.4.1 Problem Definition	21
3.4.2 Stratified Train–Test Split	21
3.4.3 Feature Standardisation	22
3.4.4 Probabilistic Prediction and Decision Threshold	22
3.4.5 Confusion Matrix and Error Types	22
3.4.6 Evaluation Metrics	22
3.4.7 Cross-validation and Out-of-Fold Estimation	23
3.5 Exploratory Data Analysis	23
3.6 Data Preprocessing	24
3.7 Data Modeling Procedure	26
3.8 SHAP Based Model Interpretation	29
CHAPTER IV RESULT & FINDINGS	30
4.1 Experiment Setup	30
4.2 Experimental Result	34
4.3 Evaluation & Findings	35
4.3.1 Generalisation Diagnostic Learning Curve Analysis	36
4.3.2 Confusion Matrix & Misclassification Patterns Error Analysis	38
4.3.3 Global Threshold Calibration using OOF Probabilities	40
4.3.4 Why XGBoost Performs Best	41
4.3.5 Findings and Implications for Model Interpretation	42
4.4 SHAP Based Interpretation	43
4.4.1 SHAP Implementation	43
4.4.2 Global Interpretation: SHAP Summary and Feature Ranking	43
4.4.3 Feature Behaviour and Non-linearity Dependence Patterns	45

4.4.4 Local Interpretation Explaining an Individual Prediction.....	47
CHAPTER V CONCLUSION & FUTURE STUDY	48
5.1 Conclusion	48
5.2 Limitations	48
5.3 Future Study.....	48
REFERENCES	49
APPENDIX	52
Appendix 1. Assistance Card.....	52
Appendix 2. Journal Article Manuscript Attachment.....	53
Appendix 4. Curriculum Vitae	54
Appendix 4. IBM Certificate	55
Appendix 5. Lecturer Assessor Revision Form	56



TABLE OF FIGURES

Figure 3. 1 Flowchart of The Research Procedure	18
Figure 3. 2 Label Quality Distribution	24
Figure 3. 3 Binary Target & Train vs Test Set Distribution.....	25
Figure 3. 4 Physicochemical Feature Correlation Heatmap	26
Figure 3. 5 Preprocessed Data	27
Figure 3. 6 Modeling Stages.....	28
Figure 3. 7 Classification Model Metrics	28
Figure 3. 8 SHAP Value	29
Figure 4. 1 Distribution of Quality Label.....	31
Figure 4. 2 Train vs Test Set Class Distribution & Alcohol Before and After Standardisation	32
Figure 4. 3 Best Confusion Matrix Model XGB	33
Figure 4. 4 F1-score Candidate Models on Test Set	35
Figure 4. 5 Evaluation Metrics of Each Model.....	36
Figure 4. 6 Learning Curve of XGBoost With Early Stopping & Threshold Tuning (Stabilised).....	37
Figure 4. 7 Confusion Matrix (Counts).....	38
Figure 4. 8 Confusion Matrix (Normalised by True Class).....	39
Figure 4. 9 Error Breakdown (FP vs FN)	40
Figure 4. 10 Threshold vs F1-score Using Out-of-Fold (OOF) Probabilities from Stratified Cross-validation	40
Figure 4. 11 Confusion Matrix for XGBoost on The Test Set Using The 0.45 Global Threshold.....	41
Figure 4. 12 SHAP summary (Beeswarm) Plot for The XGBoost Classifier on The Test Set	44
Figure 4. 13 Global Feature Importance Based on Mean Absolute SHAP Values.	45
Figure 4. 14 SHAP Dependence Alcohol	46
Figure 4. 15 SHAP Dependence Sulphates	46
Figure 4. 16 SHAP Waterfall Local	47

TABLES

Table 2. 1 Preliminary Studies	13
Table 4. 1 Summarises of The Main Hyperparameters Used for Each Classifier	32
Table 4. 2 The Accuracy, Precision, Recall, F1-score, Obtained by Each Candidate Model on The Test Set	34

